Name: _____

GCE Stats section 7

Correlation and linear regression

practice questions from GCE Maths part 2

Date:

Question parts in red cover content in the GCE Statistics specification but outside this section of the scheme of work

Time:

Total marks available:

Total marks achieved: _____

Questions

Q21.

A mobile library has 160 books for children on its records. The librarian believes that books with fewer pages are borrowed more often. He takes a random sample of 10 books for children.

(a) Explain how the librarian should select this random sample.

(2)

The librarian ranked the 10 books according to how often they had been borrowed, with 1 for the book borrowed the most and 10 for the book borrowed the least. He also recorded the number of pages in each book. The results are in the table below.

Book	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Borrowing rank	1	2	3	4	5	6	7	8	9	10
Number of pages	50	212	115	80	301	90	356	283	152	317

(b) Calculate Spearman's rank correlation coefficient for these data.

(4)

(c) Test the librarian's belief using a 5% level of significance. State your hypotheses clearly.

(3)

(Total for question = 9 marks)

Q22.

A journalist is investigating factors which influence people when they buy a new car. One possible factor is fuel efficiency. The journalist randomly selects 8 car models. Each model's annual sales and fuel efficiency, in km/litre, are shown in the table below.

Car model	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Annual sales	1800	5400	18100	7100	9300	4800	12200	10700
Fuel efficiency	5.2	18.6	14.8	13.2	18.3	11.9	16.5	17.7

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

The journalist believes that car models with higher fuel efficiency will achieve higher sales.

(b) Stating your hypotheses clearly, test whether or not the data support the journalist's belief.
Use a 5% level of significance.

(4)

(c) State the assumption necessary for a product moment correlation coefficient to be valid in this case.

(1)

(d) The mean and median fuel efficiencies of the car models in the random sample are 14.5 km/litre and 15.65 km/litre respectively. Considering these statistics, as well as the distribution of the fuel efficiency data, state whether or not the data suggest that the assumption in part (c) might be true in this case. Give a reason for your answer. (No further calculations are required.)

(1)

(Total 11 marks)

Q23.

Interviews for a job are carried out by two managers. Candidates are given a score by each manager and the results for a random sample of 8 candidates are shown in the table below.

Candidate	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Manager *X*	62	56	87	54	65	15	12	10
Manager *Y*	54	47	71	50	49	25	30	44

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

(b) Test, at the 5% level of significance, whether there is agreement between the rankings awarded by each manager. State your hypotheses clearly.

(5)

Manager Y later discovered he had miscopied his score for candidate D and it should be 54.

(c) Without carrying out any further calculations, explain how you would calculate Spearman's rank correlation in this case.

(2)

(Total 12 marks)

Q24.

The table below shows the number of students per member of staff and the student satisfaction scores for 7 universities.

University	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Number of students per member of staff	14.2	13.1	13.3	11.7	10.5	15.9	10.8
Student satisfaction score	4.1	4.2	3.8	4.0	3.9	4.3	3.7

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

(b) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence of a correlation between the number of students per member of staff and the student satisfaction score.

(3)

(Total 8 marks)

Q25.

A county councillor is investigating the level of hardship, h, of a town and the number of calls per 100 people to the emergency services, c. He collects data for 7 randomly selected towns in the county. The results are shown in the table below.

Town	*A*	*B*	*C*	*D*	*E*	*F*	*G*
h	14	20	16	18	37	19	24
c	52	45	43	42	61	82	55

(a) Calculate the Spearman's rank correlation coefficient between h and c.

(6)

After collecting the data, the councillor thinks there is no correlation between hardship and the number of calls to the emergency services.

(b) Test, at the 5% level of significance, the councillor's claim. State your hypotheses clearly.

(4)

(Total 10 marks)

Q26.

A junior judge is being trained by a senior judge to learn how to assess ice skaters. After the training, the judges each assess 6 ice skaters *A*, *B*, *C*, *D*, *E* and *F*. They each list them in order of preference with the best ice skater first. The results are shown in the table below.

Rank	1	2	3	4	5	6
Senior Judge	*A*	*B*	*D*	*C*	*F*	*E*
Junior Judge	*B*	*D*	*A*	*F*	*C*	*E*

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

(b) Test, at the 5% level of significance, whether or not there is evidence of a positive correlation between the rankings of the junior judge and the senior judge. State your hypotheses clearly.

(4)

(c) Comment on the effectiveness of the training delivered by the senior judge.

(1)

(Total for question = 10 marks)

Q27.

(a) Describe when you would use Spearman's rank correlation coefficient rather than the product moment correlation coefficient to measure the strength of the relationship between two variables.

(1)

A shop sells sunglasses and ice cream. For one week in the summer the shopkeeper ranked the daily sales of ice cream and sunglasses. The ranks are shown in the table below.

	Sun	Mon	Tues	Weds	Thurs	Fri	Sat
Ice cream	6	4	7	5	3	2	1
Sunglasses	6	5	7	2	3	4	1

(b) Calculate Spearman's rank correlation coefficient for these data.

(3)

(c) Test, at the 5% level of significance, whether or not there is a positive correlation between sales of ice cream and sales of sunglasses. State your hypotheses clearly.

(4)

The shopkeeper calculates the product moment correlation coefficient from his raw data and finds $r = 0.65$

(d) Using this new coefficient, test, at the 5% level of significance, whether or not

there is a positive correlation between sales of ice cream and sales of sunglasses.

(2)

(e) Using your answers to part (c) and part (d), comment on the nature of the relationship between sales of sunglasses and sales of ice cream.

(1)

(Total for question = 11 marks)

Q28.

The table below shows the population and the number of council employees for different towns and villages.

Town or village	Population	Number of council employees
A	211	10
B	356	2
C	1047	12
D	2463	21
E	4892	16
F	6479	25
G	6571	67
H	6573	45
I	9845	48
J	14 784	34

(a) Find, to 3 decimal places, Spearman's rank correlation coefficient between the population and the number of council employees.

(5)

(b) Use your value of Spearman's rank correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level. State your hypotheses clearly.

(4)

It is suggested that a product moment correlation coefficient would be a more suitable calculation in this case. The product moment correlation coefficient for these data is 0.627 to 3 decimal places.

(c) Use the value of the product moment correlation coefficient to test for evidence of a positive correlation between the population and the number of council employees. Use a 2.5% significance level.

(2)

(d) Interpret and comment on your results from part (b) and part (c).

(2)

(Total 13 marks)

Q29.

A researcher claims that, at a river bend, the water gradually gets deeper as the distance from the inner bank increases. He measures the distance from the inner bank, b cm, and the depth of a river, s cm, at seven positions. The results are shown in the table below.

Position	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Distance from inner bank b cm	100	200	300	400	500	600	700
Depth s cm	60	75	85	76	110	120	104

(a) Calculate Spearman's rank correlation coefficient between b and s.

(6)

(b) Stating your hypotheses clearly, test whether or not the data provides support for the researcher's claim. Use a 1% level of significance.

(4)

(Total 10 marks)

Q30.

The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown in the table below

Man	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Height (x)	1.68	1.74	1.75	1.76	1.78	1.82	1.84	1.88	1.98
Weight (y)	75	76	100	77	90	95	110	96	120

(a) Given that $S_{xx} = 0.0632$, $S_{yy} = 1957.5556$ and $S_{xy} = 9.3433$ calculate, to 3 decimal places, the product moment correlation coefficient between height and weight for these men.

(2)

(b) Use your value of the product moment correlation coefficient to test whether or not there is evidence
of a positive correlation between the height and weight of men. Use a 5 % significance level.
State your hypotheses clearly.

(4)

Peter does not know the heights or weights of the 9 men. He is given photographs of them and asked to put them in order of increasing weight. He puts them in the order

A C E B G D I F H

(c) Find, to 3 decimal places, Spearman's rank correlation coefficient between Peter's order and the
actual order.

(6)

(d) Use your value of Spearman's rank correlation coefficient to test for evidence of Peter's ability to correctly order men, by their weight, from their photographs. Use a 5% significance level and state your hypotheses clearly.

(4)

(Total 16 marks)

Examiner's Report

Q21.

A straightforward starter question dealing with correlation with most students managing the coefficient easily. They carried out the test well and typically gave its result in context.

The sampling only caused problems to those who are content with text book answers or who decided to go down the systematic route.

Q22.

This was a productive start to the paper for the majority of students with completely correct responses seen often in question (a). The hypothesis test in question (b) usually had the correct critical value, but common errors were the hypotheses being badly formed in words. Weaker responses usually missed the final conclusion in context. Question (c) was known to more students than might have been expected and if the correct terms were not known in question (d), a justification by describing skewness was not uncommon.

Q23.

This was a good start to the paper for the majority of candidates with part (a) usually being awarded full marks. In part (b) some candidates chose a two-tailed test when a one-tailed test was expected. There were very few attempts that scored both marks in part (c) with the majority being unaware of the need to use the product moment correlation coefficient and some deciding to simply ignore the tied ranks.

Q24.

The calculation of rs was, as usual, completed very well with only small minorities making slips in ranking or forgetting the "1 – " when evaluating their answer. There were a good number of fully correct responses to part (b) too although some still fail to state their hypotheses in terms of ρ. A number of candidates did not ensure that their critical value matched their alternative hypothesis and of course there were the usual crop of conclusions that did not include a reference to the context.

Q25.

This question was answered very well by the majority of candidates, with all but the very weakest candidates offering complete solutions. There was the occasional solution that gave hypotheses only in words, but this was unusual.

Q26.

Most students found part (a) straightforward scoring full marks with arithmetic slips or incorrect use of formula rarely seen. Pleasingly, reversed ranks were rarely seen. The most common error from a number of students was to 'code' the letters (A=1, B=2, etc.) and use these as ranks, leading to a wholly incorrect solution.

In part (b), there were many good answers here with most students able to identify the correct critical value from the tables and compare with their value to reject the null hypothesis. Most had correct hypotheses with very few in terms of r or simply stating them in words. The most common mark to be missed was for the final conclusion lacking some context e.g. no references to Judges, or referencing correlation but omitting 'positive'.

Part (c) was usually correct stating in various ways that the training was not effective.

Q27.

Many students are able to offer a correct reason in part (a), usually referencing normality or less often linearity. Some answers incorrectly referred to tied ranks or vaguely to discrete or subjective data. Calculations in part (d) were correct most of the time. In part (c), hypotheses were usually stated correctly in terms of ρ rather than r although a few erroneously stated in words. The correct critical values were usually found in parts (c) and (d) and most explicitly stated a correct decision for the null hypothesis. However, the mark most commonly lost was for concluding in context; too many were incomplete referring only to correlation rather than positive correlation. A majority of students realised that the difference in the results from part (c) and part (d) meant there was a non-linear relationship. There are a number, however, who appear unaware of the distinction between the two coefficients.

Q28.

Part (a) was answered very well and most scored full marks too for the hypothesis test in part (b) but a few failed to give their hypotheses in terms of ρ. Part (c) was answered well but then in part (d) few knew how to interpret these results. A handful realized that the product moment correlation coefficient (pmcc) was measuring the degree of a linear relationship and because Spearman's used ranks a non-linear relationship between the variables could be present. A few more mentioned that a bivariate normal distribution is needed to use pmcc but this assumption is not required for Spearman's.

Q29.

Most candidates secured full marks in part (a) with only a small number making arithmetic errors and a tiny minority failing to use ranks or using an incorrect formula.

The hypothesis test in part (b) was often answered very well too. Some did not use ρ for the hypotheses and some failed to give a full conclusion in context but most had the correct critical value and gave a correct statement about H_0.

Q30.

Question (a) was well done by the great majority but a mark was often lost carelessly by not giving the answer as required. The test in question (b) was usually done well with most students speaking of positive correlation in their answer. Ranking accurately was an issue in question (c), with many ranking height rather than weight, but the required technique was clearly evident. The test in question (d) was also well done by most students with good answers in context using key words. There were more two tailed tests attempted here and a failure to relate their result to the question was more evident that earlier in the question.

Mark Scheme

Section 7	Correlation and linear regression part 2
qu.21	May 2015 S3 - Q1bc
qu.22	May 2014 S3R - Q1
qu.23	June 2012 S3 - Q1
qu.24	June 2013 S3 - Q2
qu.25	June 2011 S3 - Q2
qu.26	May 2017 S3 - Q3 all (c*)
qu.27	May 2016 S3 - Q3
qu.28	June 2013 S3R - Q3
qu.29	June 2010 S3 - Q4
qu.30	May 2014 S3 - Q8

Q21.

Question Number	Scheme	Marks
(a)	Label all the books from 1 – 160 (o.e.) Use random numbers to select the 10 books	B1 B1 (2)
(b)	<table><tr><td>Book</td><td>*A*</td><td>*B*</td><td>*C*</td><td>*D*</td><td>*E*</td><td>*F*</td><td>*G*</td><td>*H*</td><td>*I*</td><td>*J*</td></tr><tr><td>Borrow rank</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Page rank</td><td>1</td><td>6</td><td>4</td><td>2</td><td>8</td><td>3</td><td>10</td><td>7</td><td>5</td><td>9</td></tr><tr><td>d^2</td><td>0</td><td>16</td><td>1</td><td>4</td><td>9</td><td>9</td><td>9</td><td>1</td><td>16</td><td>1</td></tr></table> $r_s = 1 - \frac{6\times 66}{10(100-1)}$ $[= 1 - 0.4]$ $= 0.6$ <u>0.6</u>	M1 M1 M1,A1 (4)
(c)	H_0: $\rho = 0$ H_1: $\rho > 0$ Critical value is 0.5636 0.6 > cv so significant result and sufficient evidence to reject H_0 There is support for the librarian's belief <u>or</u> there is evidence of a correlation between the number of pages in a book and the number of times it is borrowed.	B1 B1 B1ft (3) **Total 9**

	Notes
(a)	1st B1 for labelling\numbering\listing\using sampling frame of all 160 books 2nd B1 for use of random numbers\selection and mentioning the number 10
(b)	1st M1 for an attempt to rank the number of pages (at least 4 correct) Allow reverse ranks 2nd M1 for attempt at d^2 row (may be implied by sight of $\sum d^2 = 66$ or 264 for reverse ranks) 3rd M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated If answer is not correct, a correct expression is required. A1 for 0.6 (or -0.6 for reverse ranks)
(c)	1st B1 for both hypotheses in terms of ρ, one tail H_1 (compatible with ranks) Allow use of ρ_s Hypotheses just in words e.g. "no correlation" score B0. 2nd B1 for cv of 0.5636 [If they have a two tail H_1 then allow 0.6485] Allow $\pm$ for reverse ranking but must be same sign as r_s If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible. 3rd B1ft for a correct contextualised comment. Must mention "librarian" (or he) <u>or</u> "number of pages" and "borrowing" Follow through their r_s and their cv (provided it is |cv| <1) Don't insist on the word "positive" or "negative" for a one-tailed test Use of "association" is B0 Independent of 1st B1 so if $\|r_s\| >$ |cv| must say there is sufficient evidence of …….(o.e.) and if $\|r_s\| <$ |cv| must say insufficient evidence of … (o.e.) regardless of their hypotheses

Q22.

Question Number	Scheme	Marks
a)	<table><tr><td>Car model</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td></tr><tr><td>Sales rank</td><td>8</td><td>6</td><td>1</td><td>5</td><td>4</td><td>7</td><td>2</td><td>3</td></tr><tr><td>Fuel efficiency rank</td><td>8</td><td>1</td><td>5</td><td>6</td><td>2</td><td>7</td><td>4</td><td>3</td></tr><tr><td>d^2</td><td>0</td><td>25</td><td>16</td><td>1</td><td>4</td><td>0</td><td>4</td><td>0</td></tr></table> M1 $\sum d^2 = 50$ M1A1 $r_s = 1 - \frac{6\sum d^2}{8(64-1)} = 1 - \frac{6 \times 50}{8 \times 63}$ M1 $r_s = \frac{204}{504} = 0.40476\ldots$ A1 awrt 0.405	(5)
b)	$H_0: \rho_s = 0$ $H_1: \rho_s > 0$ (accept ρ_s or ρ) B1 1 tail critical value $\rho = 0.6429$ B1 Test value is not in critical region so insufficient evidence to reject H_0 M1A1ft No significant evidence at 5% level to support journalist's belief	(4)
c)	Underlying (bivariate) **Normal** distribution B1	(1)
d)	Evidence does not support Normal distribution since mean< median or (negative) skew, oe B1	(1)
		(11 marks)

Notes	
a)	M1 for attempting to rank at least one set of data A1 for at least one set of data ranked correctly(NB this mark comes after 2nd M1) M1 for attempting Σd^2 M1 for correct use of formula for r_s
b)	B1 for H_0 **and** H_1 correct (condone $\leq$ for H_0) 2nd B1 allow 0.7381 if their H_1: $\rho_s \neq 0$ M1 for correct statement relating their test statistic and critical value A1ft their test statistic, H_1 and critical value but must be in context.
c)	B1 require **Normal** distribution, ignore additional assumptions
d)	B1 require not Normal **and** valid reason

Q23.

Question Number	Scheme	Marks
(a)		

X	Y	Rank X	Rank Y	d	d^2
62	54	3	2	1	1
56	47	4	5	-1	1
87	71	1	1	0	0
54	50	5	3	2	4
65	49	2	4	-2	4
15	25	6	8	-2	4
12	30	7	7	0	0
10	44	8	6	2	4

M1
M1

$\sum d^2 = 18$ A1

$r_s = 1 - \frac{6\times 18}{8(64-1)} = 0.7857...$ awrt 0.786 M1A1

(5)

(b)

$H_0 : \rho = 0$ B1

$H_0 : \rho > 0$ B1

Critical region $r_s > 0.6429$ B1

(0.7857>0.6429 sufficient evidence to) reject H_0 M1

There is evidence of agreement between the scores awarded by each manager A1ft

(5)

(c)

(A and D are now) tied ranks (for Manager Y) B1

Average rank (awarded to A and D) **and** use $r_s = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ B1

(2)

Total 12

Notes

(a)

1st M1 for an attempt to rank score X and score Y

2nd M1 for attempting d^2 for their ranks. Must be using ranks.

1st A1 for sum of 18

3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.

2nd A1 for awrt 0.786

(b)

1st B1 for null hypotheses in terms of ρ or ρ_s

2nd B1 for alt hyp as given

3rd B1 for cv of +0.6429 (or 0.7381 if two tailed from hyp)

M1 for a correct statement relating their r_s with their cv but cv must be such that |cv|<1

A1ft for a correct contextualised comment. Must mention "scores / rankings" and "manager"

Follow through their r_s and their cv (provided it is |cv| <1

Use of "association" is A0

(c)

1st B1 Tied ranks can be implied by 2.5, 6.5 or **both** 2 or 6 or description.

2nd B1 Average rank implied by 2.5 or 6.5 or description and 'use of pmcc'.

Q24.

Question Number	Scheme	Marks
(a)	<table><tr><td>Uni</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td></td></tr><tr><td>Staff-Stu</td><td>2</td><td>4</td><td>3</td><td>5</td><td>7</td><td>1</td><td>6</td><td></td></tr><tr><td>Satisfaction</td><td>3</td><td>2</td><td>6</td><td>4</td><td>5</td><td>1</td><td>7</td><td></td></tr><tr><td>[d]</td><td>−1</td><td>2</td><td>−3</td><td>1</td><td>2</td><td>0</td><td>−1</td><td></td></tr><tr><td>d^2</td><td>1</td><td>4</td><td>9</td><td>1</td><td>4</td><td>0</td><td>1</td><td>20</td></tr></table>	M1A1A1
	$r_s = 1 - \frac{6\times 20}{7(49-1)} = 0.642857....$ (accept $\frac{9}{14}$) (awrt **0.643**)	dM1A1
		(5)
(b)	H_0: $\rho = 0$	
	H_1: $\rho \neq 0$ ($\rho > 0$)	B1
	Critical value is ± 0.7857 (± 0.7143 for a one tailed test)	B1
	0.643<cv so insufficient evidence to reject H_0 There is insufficient evidence to suggest a (positive) correlation between staff-student ratio and satisfaction.	B1ft
		(3)
		Total 8

Notes

(a)

1st M1 for an attempt to rank the staff-students ratio or satisfaction (at least 4 correct)

1st A1 for correct rankings for both (one or both may be reversed)

2nd A1 for $\sum d^2 = 20$ or correct d^2 row (NB $\sum d^2 = 92$ for one set of reversed ranks)

2nd dM1 for use of the correct formula, follow through their $\sum d^2$ (Dependent on 1st M1)

If answer is not correct, a correct expression is required.

3rd A1 If $\sum d^2 = 20$ for awrt 0.643 or if $\sum d^2 = 92$ for awrt – 0.643 (accept $\pm\frac{9}{14}$)

(b)

1st B1 for both hypotheses in terms of ρ, one tail H_1 must be compatible with their ranking

Hypotheses just in words e.g. "no correlation" score B0

2nd B1 for cv of 0.7857 or 0.7143 for one-tailed test (accept ±)

Their cv must be compatible with their H_1 which may be in words

If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible.

3rd B1ft for a correct contextualised comment. Must mention "ratio" or "no. of students per member of staff" and "satisfaction"

Follow through their r_s and their cv (provided it is |cv| <1)

Don't insist on the word "positive" for a one-tailed test

Use of "association" is B0

Independent of 1st B1 so if $|r_s| > |cv|$ must say there is sufficient evidence of …….(o.e.)

and if $|r_s| < |cv|$ must say insufficient evidence of … (o.e.) regardless of their hypotheses

Contradictory statements score B0

(This mark is just testing interpretation of comparison of their r_s and their cv)

Q25.

Question Number	Scheme	Marks
(a)	<table><tr><td>Town</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td></tr><tr><td>h rank</td><td>1</td><td>5</td><td>2</td><td>3</td><td>7</td><td>4</td><td>6</td></tr><tr><td>c rank</td><td>4</td><td>3</td><td>2</td><td>1</td><td>6</td><td>7</td><td>5</td></tr><tr><td>$\lvert d\rvert$</td><td>3</td><td>2</td><td>0</td><td>2</td><td>1</td><td>3</td><td>1</td></tr><tr><td>d^2</td><td>9</td><td>4</td><td>0</td><td>4</td><td>1</td><td>9</td><td>1</td></tr></table> M1 M1 $\sum d^2 = 28$ M1A1 $r_s = 1 - \frac{6\times 28}{7\times 48}$ M1 $= 0.5$ A1 (6)	
(b)	$H_0 : \rho = 0, H_1 : \rho \neq 0$ B1 Critical values are $r_s = \pm 0.7857$ B1ft 0.5<0.7857 insufficient evidence to reject H_0 M1 Councillor's claim is supported. A1ft (4) **10**	
Notes (a)	1^{st} M1 for an attempt to rank the hardship against calls 2^{nd} M1 for attempting d for their ranks. Must be using ranks. 3^{rd} M1 for attempting $\sum d^2$ (must be using ranks) 1^{st} A1 for sum of 28 (or 84) 4^{th} M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required. 2^{nd} A1 for awrt 0.5 (or -0.5)	
(b)	1^{st} B1 for both hypotheses in terms of ρ, H_1 must be two tail. 2^{nd} B1 for cv of $\pm$0.7857 (or 0.7143 to ft from 1-tailed H_1) M1 for a correct statement relating their r_s with their cv but cv must be such that $\lvert cv\rvert < 1$ A1ft for a correct contextualised comment. Must mention "Councillor" and "claim" or "hardship" and "number of calls (to the emergency services)" Follow through their r_s and their cv (provided it is $\lvert cv\rvert < 1$ Condone use of "association" in conclusion for A1 Condone 'positive' in conclusion.	

Q26.

Question Number	Scheme	Notes	Marks
(a)	<table><tr><th></th><th>Senior Judge</th><th>Junior Judge</th><th>d</th><th>d^2</th></tr><tr><td>A</td><td>1</td><td>3</td><td>-2</td><td>4</td></tr><tr><td>B</td><td>2</td><td>1</td><td>1</td><td>1</td></tr><tr><td>C</td><td>4</td><td>5</td><td>-1</td><td>1</td></tr><tr><td>D</td><td>3</td><td>2</td><td>1</td><td>1</td></tr><tr><td>E</td><td>6</td><td>6</td><td>0</td><td>0</td></tr><tr><td>F</td><td>5</td><td>4</td><td>1</td><td>1</td></tr><tr><td></td><td></td><td></td><td></td><td>8</td></tr></table>	M1 for an attempt to rank judges lists (at least 4 correct for each judge) A1 for correct rankings for both (may be reversed). Can be implied by correct d^2 or r_s N.B. Table could be ordered in terms of Senior Judge.	M1A1
	$\sum d^2 = 8$ or 62	8 or 62 or correct d^2 row	A1
	$r_s = 1 - \frac{6\times 8}{6\times 35} = \frac{27}{35} = 0.771$	M1 for use of the correct formula, follow through their $\sum d^2$ (Dependent on 1st M1) If answer is not correct, a correct expression is required. A1 exact fraction or awrt $(\pm)0.771$	dM1A1
			(5)
(b)	$H_0 : \rho = 0$ $H_1 : \rho > 0$	Both hypotheses in terms of ρ or ρ_s . Hypotheses just in words e.g. "no correlation" score B0	B1
	Critical value ± 0.8286	Accept ± 0.8857 if 2-tailed H_1.	B1
	(0.771<0.8286) so insufficient evidence to reject H_0	Follow through their r_s and their c.v. if |cv| <1 and $\lvert r_s \rvert < 1$	M1
	There is insufficient evidence to suggest a **positive** correlation between the judges.	A correct contextualised comment that includes "judges".	A1 ft
			(4)
(c)	(For positive correlation c.v.is 0.8286>0.771)		
	Training of junior judge was ineffective.	o.e. Follow through from their cv and r_s	B1 ft
			(1)
			Total 10

Q27.

Question Number	Scheme	Marks
(a)	Variables do not have a (joint) normal distribution Relationship is not linear The given data is ordinal Any 1	B1 (1)
(b)	<table><tr><td>Day</td><td>Sun</td><td>Mon</td><td>Tues</td><td>Weds</td><td>Thurs</td><td>Fri</td><td>Sat</td></tr><tr><td>Ice-cream</td><td>6</td><td>4</td><td>7</td><td>5</td><td>3</td><td>2</td><td>1</td></tr><tr><td>Sunglasses rank</td><td>6</td><td>5</td><td>7</td><td>2</td><td>3</td><td>4</td><td>1</td></tr><tr><td>d</td><td>0</td><td>-1</td><td>0</td><td>3</td><td>0</td><td>-2</td><td>0</td></tr><tr><td>d^2</td><td>0</td><td>1</td><td>0</td><td>9</td><td>0</td><td>4</td><td>0</td></tr></table> $\sum d^2 = 14$ $r_s = 1 - \frac{6\times 14}{7(49-1)} = 0.75$	M1 M1A1 (3)
(c)	$H_0 : \rho = 0$, $H_1 : \rho > 0$ 5% cv 0.7143 Reject H_0 Evidence of **positive** correlation between sales of **ice cream** and sales of **sunglasses.**	B1 B1 M1 A1cao (4)
(d)	(5% cv 0.6694) Accept H_0 Insufficient evidence of **positive** correlation between sales of ice cream and sales of sunglasses.	M1 A1cao (2)
(e)	Suggests relationship might be non-linear.	B1 (1) **Total 11**

Notes	
(a)	Accept 'already ranked' oe for ordinal Accept one variable is not normally distributed
(b)	M1 attempt to find d, d^2 and sum. may be implied by sight of $\sum d^2 = 14$ M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated. If answer is not correct, a correct expression is required. A1 0.75 cao
(c)	1st B1for both hypotheses in terms of ρ, one tail. Allow use of ρ_s. **Only award** if no errors seen in hypotheses in part(c) **and part(d)** Hypotheses just in words e.g. "no correlation" score B0. B1 0.7143 cao M1 must correctly reject / not reject the null hypothesis based on their test stat and cv oe A1 Conclusion must mention ice cream and sunglasses
(d)	M1 for not rejecting / accepting null hyp A1 must mention ice cream and sunglasses

Q28.

Question Number	Scheme	Marks
(a)	<table><tr><td>Town</td><td>*A*</td><td>*B*</td><td>*C*</td><td>*D*</td><td>*E*</td><td>*F*</td><td>*G*</td><td>*H*</td><td>*I*</td><td>*J*</td></tr><tr><td>Pop</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Empl</td><td>2</td><td>1</td><td>3</td><td>5</td><td>4</td><td>6</td><td>10</td><td>8</td><td>9</td><td>7</td></tr><tr><td>$\lvert d\rvert$</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>3</td><td>0</td><td>0</td><td>3</td></tr><tr><td>d^2</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>9</td><td>0</td><td>0</td><td>9</td></tr></table>	M1
	$\sum d^2 = 22$	M1A1
	$r_s = 1 - \frac{6 \times 22}{10 \times 99}$	dM1
	$= \frac{143}{165} = 0.866\dot{}$ awrt **0.867**	A1 (5)
(b)	$H_0: \rho = 0$; $H_1: \rho > 0$	B1
	CV = 0.6485	B1
	in critical region / significant/ reject H_0	M1
	evidence of positive correlation between population and no. of employees	A1 (4)
(c)	CV = 0.6319	B1
	[not in critical region / not significant/ do not reject H_0]	
	No evidence of positive correlation	B1 (2)
(d)	No evidence to suggest that as pop' increased the no. of employees increased linearly.	B1
	Villages ranked highly for pop' were also ranked highly for the no. of employees.	B1 (2)
ALT	Alternate for part (d) if different conclusions in part (b) and part (c)	
	Data probably not (bivariate) normal therefore Spearman's coefficient is more suitable than the product moment correlation coefficient.	**[Total 13]**

Notes

(a)

1st M1 for an attempt to rank no of employees against the populations

2nd M1 for attempting $\sum d^2$ (must be using ranks) ft their ranks

1st A1 for 22

3rd dM1 dep on 1st M1 for use of the correct formula with their $\sum d^2$.

If ans. is not correct an expr' is required.

(b)

1st B1 for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking

M1 for a correct statement relating their r_s ($|r_s| < 1$)with their cv but cv must be such that |cv|<1

A1 for a correct contextualised comment that is rejecting H_0

Must mention "population" and "no. of employees" and "positive correlation".

Follow through their r_s and their cv (provided it is |cv| <1 Use of "association" is A0

(c)

1st B1 for 0.6319 2nd B1 does not require context just no positive correlation mentioned

(d)

1st B1 for a comment relating to pmcc

(i) no linear relationship or (ii) pmcc requires (joint) normal distribution

2nd B1 for a second comment relating to Spearman's

(i) there is a (non-linear) relationship between ranks..

or (ii) data not (joint) normal so Spearman's is better

Q29.

Question Number	Scheme	Marks
(a)	<table><tr><td>Distance rank</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>Depth rank</td><td>1</td><td>2</td><td>4</td><td>3</td><td>6</td><td>7</td><td>5</td></tr><tr><td>$\lvert d\rvert$</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>d^2</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>4</td></tr></table>	M1 M1
	$\sum d^2 = 8$	M1A1
	$r_s = 1-\frac{6\times 8}{7\times 48}$	M1
	$=\frac{6}{7}=0.\dot{8}57142\dot{}$ awrt **0.857**	A1 (6)
(b)	$H_0 : \rho = 0, H_1 : \rho > 0$	B1
	Critical value at 1% level is 0.8929	B1
	$r_s < 0.8929$ so not significant evidence to reject H_0,	M1
	The researcher's claim is not correct (at 1% level). or insufficient evidence for researcher's claim or there is insufficient evidence that water gets deeper further from inner bank. or no (positive) correlation between depth of water and distance from inner bank	A1ft (4) **10**

(a)	1st M1	for an attempt to rank the depths against the distances
	2nd M1	for attempting d for their ranks. Must be using ranks.
	3rd M1	for attempting $\sum d^2$ (must be using ranks)
	1st A1	for sum of 8 (or 104 for reverse ranking)
	4th M1	for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.
	2nd A1	for awrt (±) 0.857. Sign should correspond to ranking (so use of 104 should get -0.857)
(b)	1st B1	for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking
	2nd B1	for cv of 0.8929 (accept ±)
	M1	for a correct statement relating their r_s with their cv but cv must be such that |cv|<1
	A1ft	for a correct contextualised comment. Must mention "researcher" and "claim" or "distance (from bank)" and "depth (of water)" Follow through their r_s and their cv (provided it is |cv| <1) Use of "association" is A0

Q30.

Question Number	Scheme	Marks
(a)	$r = \frac{9.3433}{\sqrt{0.0632 \times 1957.5556}}$ $= 0.840$	M1 A1 (2)
(b)	$H_0 : \rho = 0$ $H_1 : \rho > 0$ Critical value = 0.5822 $0.840 > 0.5822$ There is evidence to reject H_0. There is evidence of a positive correlation between a man's height and his weight.	B1 B1 M1 A1ft (4)
(c)		

Man	A	B	C	D	E	F	G	H	I
Actual weight	1	2	7	3	4	5	8	6	9
Peter's order	1	4	2	6	3	8	5	9	7
d^2	0	4	25	9	1	9	9	9	4

	B1 B1
$\sum d^2 = 70$	M1A1
$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ $= 1 - \frac{6\times 70}{9(81-1)}$ $= 0.417$	dM1 A1 (6)
(d) $H_0 : \rho = 0$ $H_1 : \rho > 0$ Critical value 0.600 $(0.417 < 0.600)$ There is insufficient evidence to reject H_0. Peter does not have the ability to correctly order men, by weight, from their photograph.	B1 B1 M1 A1 (4) **Total 16**

Notes		
(a)	M1 Clear use of $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ A1 0.840 cao	
(b)	1st B1 for both hypotheses in terms of ρ , one tail H_1 must be compatible with their r Hypotheses just in words e.g. "no correlation" score B0 2nd B1 for 0.5822 cao M1 for a statement comparing 'their r' with 'their cv' A1 for a correct contextualised comment. Must mention positive correlation, be carrying out a 1-tailed test and mention height and weight. Follow through their r and their cv (provided their $\|cv\| < 1$ and their $\|r\| < 1$)	
(c)	1st B1 for attempt to rank actual weight / Peter's order with at least 4 correct 2nd B1 for correct rankings for both (one or both may be reversed) 1st M1 for use of $\sum d^2$ with at least 4 values correct and attempt to add 1st A1 for 70 or 170 with reversed rankings 2nd dM1 for use of the correct formula, follow through their $\sum d^2$ .Dependent on 1st M1 If answer is not correct, a correct expression is required. 2nd A1 for awrt 0.417 or $\frac{5}{12}$	
(d)	1st B1 for both hypotheses in terms of ρ or ρ_s One tail H_1 must be compatible with their ranking Hypotheses just in words e.g. "no correlation" score B0 2nd B1 for cv of 0.6(00) cao Their cv must be compatible with their H_1 which may be in words M1 for statement comparing 'their r' with 'their cv' A1 for a correct contextualised comment. Must mention Peter and Men. Follow through their r and their cv (provided their $\|cv\| < 1$ and their $\|r_s\| < 1$)	